DOCUMENT RESUME

ED 418 660 HE 031 189

AUTHOR Yorke, Mantz
TITLE Can Performance Indicators Be Trusted?
PUB DATE 1997-05-00
NOTE 43p.; Revised version of a paper presented at the AIR Forum (37th, Orlando, FL, May 18-21, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Policy; Foreign Countries; Higher Education; *Institutional Evaluation; *Performance Based Assessment; Politics of Education; *Reliability; *Validity
IDENTIFIERS Performance Funding; *Performance Indicators; United Kingdom

ABSTRACT

This paper examines the purpose, validity, and reliability of performance indicators in higher education, focusing on the experience of higher education institutions in the United Kingdom. Specifically, it studies the role of student entry and exit performance, teaching and staff quality, retention and completion, and placement in employment as indicators of higher education performance. It examines these indicators in terms of the interests they serve, their validity, and their robustness (including their vulnerability to manipulation). The paper discusses the problems associated with each set of indicators, noting that the validity and reliability of such indicators are mixed and that government attempts to tie funding to performance indicators may have unintended consequences. The paper concludes that the belief that performance indicators can provide accurate information on institutional functioning appears to have weakened in recent years, with early optimism having becoming tempered by debate and experience. (Contains 40 references.) (MDM)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

HE

# Can performance indicators be trusted?

ED 418 660

**Mantz Yorke**

Director

Centre for Higher Education Development

Liverpool John Moores University

IM Marsh Campus

Barkhill Road

Liverpool

L17 6BD

England

E-mail m.yorke@livjm.ac.uk

Revised version of a paper presented at the 37th AIR Forum, Orlando, 18-21 May 1997.

HE 031 189

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Mantz Yorke

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

# Can performance indicators be trusted?

**Abstract**

Performance indicators are well-established in the language of accountability in higher education, and are used to serve a variety of political and micro-political ends. The speed of their implementation, however, has not been matched by equivalent progress in the development of their technical qualities. This article subjects a selection of indicators to analysis in terms of (a) the interests they serve, (b) their validity and (c) their robustness (including their vulnerability to manipulation). In the light of this analysis, some comments are made regarding the use of indicators in policy arenas.

**Introduction**

Performance indicators are now standard components of the 'language' of management in, and of, education around the world. In part, their development has been fuelled by an increasing concern on the part of governments at national/state level to assure themselves that the education service (to which they contribute funding) is delivering what they see as being desirable, and at a cost that can be afforded. The use of performance indicators has also been given an impetus, particularly in the United Kingdom, by a scepticism on the part of government that, left alone to deliver the service expected of them, professional groups cannot entirely be trusted. Institutions have always espoused a commitment to the development of students but, until comparatively recently, have perhaps been less assiduous than they might have been in demonstrating their successes in this respect. The emphasis on outcomes assessment in the United States, and that on the notion of 'graduateness' (HEQC, 1996a) in the United Kingdom, are indicative of governmental concern that higher education should not only serve national or state goals, but also that this service should be clearly demonstrated.

However, it is not only governments which are interested in performance indicators: there is a wide range of stakeholders for whom institutional performance is important. Governments appear at one end, but intending students (and their sponsors) appear at the other, with other organisations such as employers and professional bodies somewhere in between.

As a result of the range of interests that are being brought to bear on the performance of the education system, performance indicators cannot be construed in 'value-neutral' terms, or as mere management statistics. They exist in political arenas of varying levels of inclusivity (e.g. federal, state, institutional) and may be used for purposes for which they were not designed: for this reason (amongst others), the interpretation of a performance indicator is very much open to contest. As an example of contestability, an institution congratulating itself for having improved the results profile of its students may be challenged by politicians (operating in respect of a different kind of agenda) who assert that it has let standards fall.

For this reason, it is important that performance indicators represent accurately what they purport to represent, and that in so doing they are robust: in other words, they need to have high validity and reliability. In the context of higher education various sets of indicators have been proposed (see, e.g. Linke, 1991; Kells, 1993; CAPIOA, 1993; JPIWG, 1994a; Ruppert, 1994). However, the indicators are often underdeveloped in technical terms and implicitly reflect particular interests. An area either largely neglected, or left to dubious proxy measures, is that of the quality of the student experience ('the educational process'). This is of obvious importance for the management of institutions: however, it is of less importance to outside interests, which tend to be interested in 'the product' and pay particular attention to outcome measures such as cost-effectiveness and the capabilities of graduating students. One of the main problems for indicator systems is the relationship between indicators of process and those of product, and this is

complicated by the fact that what is subsumed by the terms 'process' and 'product' will vary with the stakeholder's perspective.

Perhaps the most thorough investigation of a set of potential performance indicators for higher education was undertaken by a committee chaired by Linke (1991) in Australia. The Linke Report's discussion of a number of indicators revealed that, whereas some lacked sufficient robustness at their current state of development, eighteen - though possessing technical weaknesses - were sufficiently well-developed to merit their application in Australia.

The state of development of performance indicators for higher education - still relatively primitive - leaves scope for considerable refinement. As a contribution to the process of refining, this paper revisits some basic concerns about indicators and their use and, in respect of a selection that bear in some way on teaching and learning, asks the following questions.

- Who wants to know what?
- For what purposes is the information to be used?
- How valid and reliable are the indicators that are being used, or are being proposed for use?
- Do the indicators have any side-effects?

The selected indicators are

- student entry and exit performance, and the related indicator of 'value added';
- teaching quality, and the related indicator of staff quality;
- retention and completion; and
- placement in employment.

In addition, some other indicators of obvious importance for institutional health are discussed much more briefly.

**A selection of performance indicators**

The selected indicators are typical of those being used, or under development, in the United Kingdom. The selection covers a spectrum from those that are fairly 'hard' to those which are distinctly 'soft'. Readers from other countries are asked to make the necessary adjustments to suit their particular situations.

Entry and exit performance, and 'value added'

Whose interest?

Government and its agencies have an obvious interest in these variables since they give an indication of the effectiveness of the higher education system. Coupling these variables with costs allows estimates to be made of the sector's efficiency and economy.

Institutions use these variables both internally and comparatively in order to assess their overall effectiveness. Some institutions will have particular reason to favour the concept of 'value added' (the extent to which students gain as a result of their time in higher education), whereas others which have high entry qualifications will disfavour such an indicator because of the 'ceiling effect'. Value added appears in some ranking tables, and therefore has some reputational value to those institutions that score highly on it.

Entry qualification expectations play a significant part in the applications process, but it is likely that exit qualifications are of relatively limited interest. Similarly, the general public is unlikely to be concerned with value added, save with respect to some general apperception of an institution helping students with relatively limited formal entry qualifications to succeed.

Validity

The validity of all three indicators is problematic, that of value added being particularly suspect. From the point of view of general governmental policy regarding a whole higher education system (with the associated issue of accountability), the roughness inherent in entry and exit qualifications may be of little importance. However, if the data are used to steer funding, then tensions arise where diversity exists in the system: institutional aims, clientèles and so on are not homogeneous. At the level of the institution it is likely that the notion of value added can be sustained since it can give an indication of whether

particular groups of students are, or are not, achieving as well as their peers. An institutional commitment to equality of opportunity demands analyses of this sort.

In the United Kingdom there is one main national examination which students use as the basis for entry to higher education (the Advanced Level of the General Certificate of Education[1]) which is set by a number of different examination boards. Recently it has been claimed that standards may not be equivalent across the boards (one board being accused of unjustifiably raising the marks in respect of some schools, perhaps because it was seeking to maintain a high pass rate for marketing purposes). The Secretary of State for Education in the previous Conservative government has made moves to reduce the number of boards offering A-level examinations as a result. Although the boards engage in comparability studies, there is no guarantee that standards are consistent across the A-level examination as a whole. The United Kingdom has no nation-wide testing service to give the equivalent of the American Scholastic Attainment Test [SAT] or American College Testing [ACT] scores for the purposes of entry to higher education.

Many institutions are prepared to accept students on the basis of their prior learning or prior experiential learning, and have devised ways in which students can demonstrate their

[1] Students may incorporate 'half A-levels' [AS-level] in their entry qualifications, and there are many other types of qualification which are taken as suitable for entry. Students in Scotland can enter higher education at 17, rather than the 18 typical of school-leavers in England, on the basis of 'Scottish Highers'.

knowledge and skills. Entry on this basis cannot easily be aligned with mainstream entry qualifications, and there is no metric along which the various types and levels of qualifying performance can be placed. At its highest, measurement on an entry scale will be ordinal - and even this will require value judgments to be made about some data which are really only categorical in character.

Exit performance is also problematic. For a long time the United Kingdom operated on the basis of the myth of the 'gold standard' of the first degree - all awards of the same classification[2] being held to be equivalent. Over the years recruitment and sponsorship practice has shown that employers and research funding bodies have not subscribed to the myth: for example, a degree from Oxford or Cambridge has tended to be a more prestigious passport than a similar class of degree from, say, a former polytechnic.

Various researches are exposing the myth of equivalence. Chapman (1994) showed in the case of geography that there was a persistent variation between departments regarding the classifications of degree awarded, and he subsequently extended his studies to cover eight subjects, where the same general conclusion held (HEQC, 1996b). Set against these findings, there has been a general rise in the level of degree awards. This is all rather difficult to disentangle, and could embody any or all of the following: genuinely better performances as curricula are nowadays designed more explicitly and delivered against

---

[2] UK bachelor's degrees with honours are usually categorised as first, upper second, lower second, and third class. 'Pass' or 'ordinary' degrees may also be awarded, but these are non-honours degrees.

intended outcomes, differing modes of assessment (typically more coursework and less formal examination), and an easing of standards. The Higher Education Quality Council (1996a) has recently reported on the need for higher education institutions to be more explicit about the standards to which they subscribe, thereby allowing programmes of study and student achievements to be compared without necessarily implying equivalence.

The Student Assessment and Classification Working Group [SACWG] has tackled other aspects of 'the comparability problem'. It demonstrated that, in six broadly similar post-1992 universities, there were persistent differences between the marks/grades given for different subjects: for example, Law typically scored low, and Mathematics, Statistics and Computing had typically high spreads of marks/grades (Yorke et al, 1996). These findings have implications for equity, particularly in institution-wide modular schemes. SACWG has also studied the decision-making algorithms for degree classifications, and has concluded that perhaps 15 per cent of students may be misclassified as a result of the nature of the algorithm and/or the way in which it is used in practice (Turner and Woolf, et al, 1997).

'Value added' is an attractive notion, and particularly so for those institutions which do not recruit the 'top quality' students. This has been an indicator favoured by the former polytechnics and the colleges since it has offered them a way of demonstrating that they could stimulate students to remarkable successes in relation to their (comparatively low) entry qualifications. The difficulty lies in actually measuring value added, since - as is

shown above - both the entry and exit indicators are problematic and, further, there is no common metric which links the two. A study by the former Council for National Academic Awards (CNAA/PCFC, 1990) showed clearly the weakness of simplistic exit/entry computations, but its preferred conclusion - that one could measure an institution's performance against national norms - also suffered from a number of technical weaknesses (see Cave et al 1997, pp127-131).

The concept of value added may be more useful in internal institutional monitoring, rather than for the purposes of inter-institutional comparison. If a matrix such as that in Figure 1 is constructed for each general programme of study, it allows an institution to see if there are any anomalies between students grouped according to entry qualification. For example, students entering without formal qualifications may do disproportionately badly compared with their peers, and this may stimulate an inquiry into the effectiveness of the chosen teaching/learning approach. Year-on-year, it may be possible to detect trends that suggest the need for action at programme level.

| EXIT QUALIFICATION<br><br>ENTRY QUALIFICATION | Fail | Pass/unclassified degree | Third class degree | Lower 2nd class degree | Upper 2nd class degree | First class degree |
|---|---|---|---|---|---|---|
| High A-level points | | | | | | |
| Medium A-level points | | | | | | |
| Low A-level points | | | | | | |
| Vocational diploma | | | | | | |
| 'Access course' qualification | | | | | | |
| ... | | | | | | |
| No formal qualifications | | | | | | |

Figure 1. A schematic representation of a value added matrix which could be used for internal institutional purposes.

Reliability

The indicators - as is apparent from the preceding discussion - decline in reliability from entry qualification, through exit qualification to value added.

Side effects

If resourcing or reputation depend to any extent on these indicators, there will be a risk that performances will evolve in the direction favourable to the institution. In the United States, Stecklow (1995) reported the inflation of SAT scores at entry by some institutions which had eliminated the scores of weak groups (such as students whose first language was not English) in order to give what they claimed was a more representative picture. As Johnes and Taylor (1990) point out, where institutions themselves control the outcomes of higher education there is an inherent temptation to move scores upwards: if the scores are in borderline areas their adjustment could be insufficiently gross to attract the attention of the external examiners used in the United Kingdom[3]. Indeed, the handling by examination boards of borderline performances is a matter worthy of study in its own right. Given that value added is a function of the other two indicators, it is potentially more open to manipulation than either.

[3] There has been some interest in the possibility of funding institutions with reference to completion rates. Since completion rates may be computed on a cross-sectional, rather than a longitudinal, basis (JPIWG, 1994a), 'upward drift' of marks/grades could take places at points when external examiners are not employed.

Teaching quality

Whose interest?

In the United Kingdom this is a matter of government interest which had its roots in the former Conservative government's commitment to consumerism: the recent change of government seems unlikely to change the position. The government interest is made manifest in the Education Act 1992, in which the funding councils for England, Scotland and Wales are required to secure that the assessment of teaching quality takes place. Further, the outcomes of such assessments are expected to inform funding decisions, although relatively little use has been made of them for this purpose.

Institutions have typically expressed a commitment to teaching quality and to its improvement. A few have pervaded their curricula with this commitment. Now that the quality of teaching is flagged in the reports of quality assessments accolades can be used in promotional activities. This has a precedent deriving from the formal inspection of the erstwhile polytechnics before they were designated as universities in 1992. Her Majesty's Inspectors were able to rate subject disciplines according to their quality, and Q and q 'flags' were applied respectively to provision whose constituent courses largely or partly met the high quality criterion. Successful institutions made much of their 'Qs' and 'qs' in advertisements.

The public, as consumers, has a clear interest in teaching quality. Reports of quality assessments are placed in the public domain, but the style in which they are written is relatively opaque. Summaries, such as those appearing in published guides to higher education, are too abbreviated to be of any real value.

Validity

The notion of 'teaching quality' has become a problematic concept with respect to a context in which student learning is to the fore. Quality assessments tend predominantly to reflect the teaching 'supply side' (such as lecture performance, laboratory and studio practice, resources, and assessment) and aspects of quality assurance, with student learning being picked up in interactive teaching sessions and in the outcomes of the higher education experience. The training of assessors is relatively short in duration, and a question exists over the extent to which assessors are able fundamentally to get to grips with teaching and learning - after all, the observation of teaching in teacher education programmes is typically given greater attention in induction into the role of a teacher educator. There is a general tendency for the higher ratings to be given to the pre-1992 universities, which are better resourced than post-1992 universities and colleges. Given that the latter were given a brief for teaching (as opposed to research), it is curious that they have done relatively less well on the teaching quality criterion. The suspicion exists that resources and reputation might be significant covariates in quality assessment.

Apart from the development of a 'student progression profile' (Williams, 1994), a report from the Scottish Centrally Funded Colleges (SCFC, 1992), and student satisfaction surveys (e.g. Ramsden, 1991a, b; Green et al, 1994) the literature on performance indicators has tended to duck the issue of teaching quality and the more general issue of the quality of the student experience (see Yorke, 1996a, for an extended discussion). Where there has been an interest in teaching quality it is often found as proxy measures such as staff quality (including the number of academics with doctorates, indicators of research activity and the like) and the quantity of resources in the library. Such proxy measures are at best indirectly related to teaching quality and, in any event, some tend to vary with the subject discipline.

Reliability

Given the problems with validity, the reliability of teaching quality assessments is unlikely to be high.

Side effects

Teaching quality assessment by the funding councils in the United Kingdom has without doubt led institutions to focus on teaching to a far greater extent than they had done before the programme was instigated. However, some institutions and commentators have claimed that the costs to the system, both direct and opportunity, have been high. This is a matter of contention. The overt costs of quality audit and quality assessment for England are of the order of £7m - about 0.1% of the total funding for higher education.

This figure, however, leaves out of consideration opportunity costs to institutions, the costs associated with internal quality assurance, and the costs of examining student achievements. In any case, given the pressures likely to obtain in the future, it can be argued that greater attention needs to be given to preparing for change (Yorke, 1996b).

There is perhaps a tendency on the part of institutions to conform to perceptions of a perceived 'ideal' approach to teaching and learning which reflects tradition, and institutions have also quickly learned how to present their self-appraisals of teaching quality in the most positive light.

One problem with the cycle of assessments is that each subject discipline is unlikely to be visited more frequently than every six years (proposals currently being worked up for a revised national system suggest an eight-year cycle [JPG, 1996]). There will inevitably be a temptation for institutions to concentrate their efforts on teaching in the period when it is under scrutiny, and to attend to different priorities (such as doing research, or generating income) at other times in the cycle.

I have argued (Yorke, 1996b) that, at a time of considerable flux for higher education, the urgent need is for the enhancement of quality rather than its assessment. If institutions were required to produce plans for improvement, and were audited on the delivery of these, then it is likely that the accountability needs could be met in passing and not need the magnitude of the current and proposed quality assessment exercises.

## Retention and completion

### Whose interest?

Government has a direct interest on the grounds of concern about the extent to which its investment is generating the desired return to the national or state economy, and an indirect interest through funding bodies which channel the national or state investment into institutions. Funding bodies might take the view that institutions should be funded on outcomes, as is the case with the Further Education Funding Council in England, rather than on the numbers of entering students.

Data on retention and completion provide an institution with evidence of the extent to which it is fulfilling its mission. Departmental data allow both internal and external benchmarking of achievements. Both levels of data are potentially useful for assisting improvement. On a more self-serving note, students are a source of institutional income, and so institutions have an obvious interest in maximising their retention rates. High retention and completion rates can also be used in self-promotion.

With higher education becoming more consumer-oriented around the world, and students in some countries bearing an increasing burden of their costs of study, data on retention and completion will increasingly be taken into account by the public. Guides such as The Times good university guide 1996 (O'Leary, 1996) include completion data, but the institution-wide nature of the index is of little practical value to the potential student seeking a place on, say a Modern Languages degree programme.

Validity

On the face of it, retention and completion statistics ought to be non-problematic: an institution knows how many students it has at any one time, and whether they succeeded in gaining the awards for which they were studying within, say, the US criterion of 150 per cent of the expected time for completion (Gaither et al, 1994, p87). These kinds of indicator ought to work fairly well for full-time programmes, but those students who draw on the possibilities inherent in credit frameworks, and those who choose part-time routes, will not fit into conceptions of progression and completion that are essentially based on full-time study.

The hypothetical data in Figure 2 show, for full-time students, how vulnerable cohort-based statistics are to misunderstanding. This three-year full-time programme accepts 100 new entrants in 1994, and also two students who are repeating the first year. At the end of the year seven exit from the programme for various reasons, and one returns to the beginning of year one (perhaps because of illness during the first year). Year 2 takes the remaining 94 students plus seven others, some coming from previous years on the programme, and others entering 'with advanced standing' from elsewhere. A similar pattern of events occurs in Year 3, from which 95 students graduate. The crude statistic based on year-by-year data would give a completion rate of 95 per cent. This clearly overestimates the true figure which, assuming that all 15 students transferring in from earlier years complete without any further delay, is 80 per cent of the 1994 new entrants.

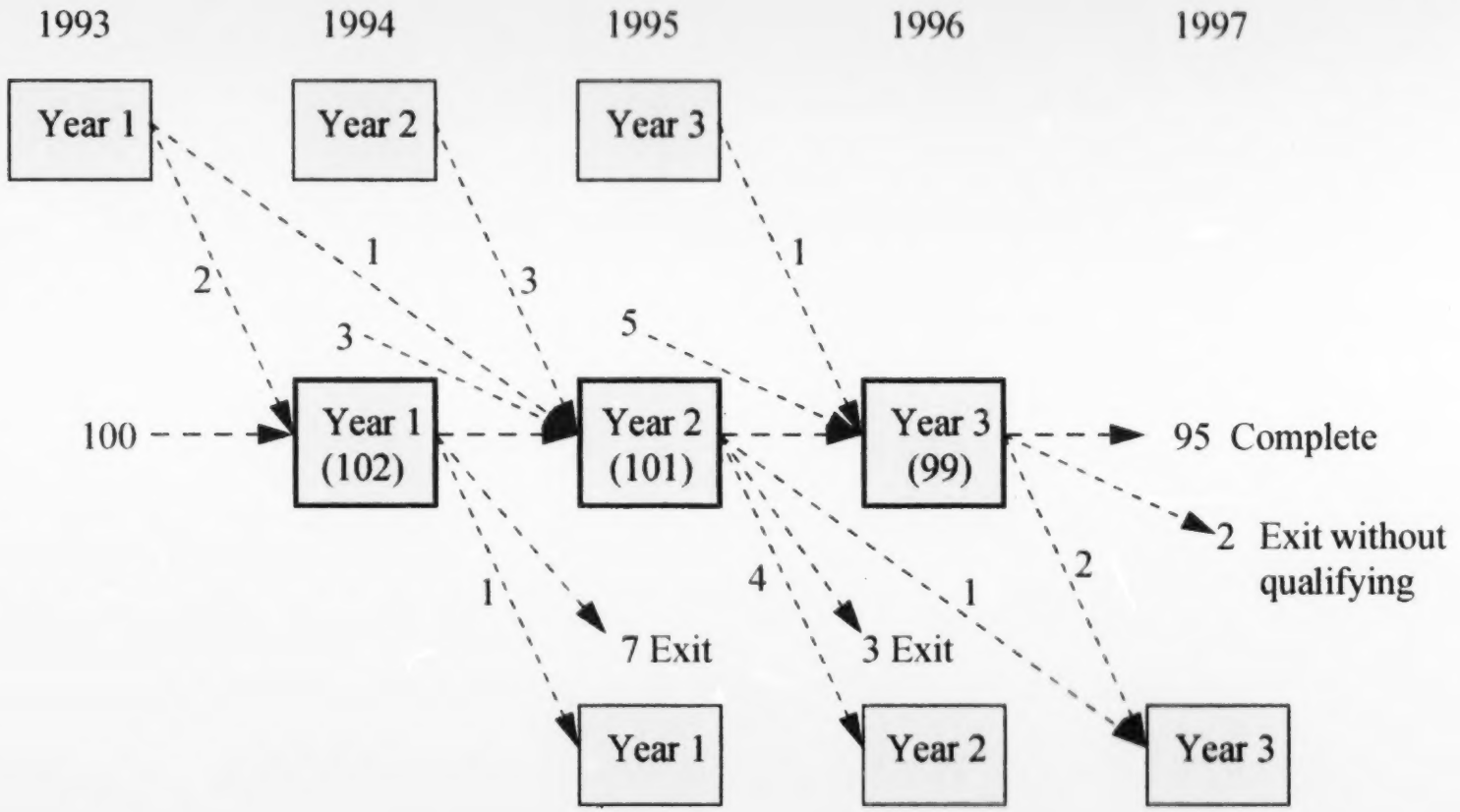


Figure 2 An illustration of the complexity of retention and completion (hypothetical data)



BEST COPY AVAILABLE

If one is trying to get an up to date performance indicator of progression, then one way of doing this is to adopt a 'census' approach in which a cross-section count of students is used (see JPIWG, 1994b for an example). With reference to Figure 2, this would mean taking a calendar year 'slice' down the page (e.g. 1995) and cumulating the year-based data. This approach exchanges precision for contemporaneity, making the assumption that success rates for a particular year of study remain much the same from year to year. It is likely to be sufficiently accurate for policy-related use at the level of the system.

Moving into unitised programmes, it is possible to compute success rates for each study unit, as the Australians do in their index 'study progress unit' or SPU (see Dobson et al, 1996). Once one gets to this level of detail, the question then arises as to how SPUs can be aggregated into yearly continuation rates and completion rates - not a trivial undertaking in the United Kingdom, where full-time attendance is slowly breaking down into something less than that because of (inter alia) the pressure on students to obtain jobs to subsidise their studies and the opportunity for greater flexibility in study offered by modularised schemes.

Reliability

The true completion rate for a programme, built up from individual students' results, is likely to be very highly reliable. Census-type indicators are likely to have a high reliability, though not as high a reliability as the true completion rate.

Side effects

However, if rewards for institutions or faculty are related to retention and completion, then the high reliability - and, by extension, validity - risks being bought at the expense of an easing of standards (and there has been a persistent concern in the United States about grade inflation, which is finding echoes in the United Kingdom regarding degree results).

If institutions are funded with reference to retention and completion, then the causes of non-completion become important. Withdrawal or non-completion may be voluntary or involuntary, and may be influenced by institutional and/or personal causes. In a study which I am currently leading, non-completion at the 'macro' level appears to be a complex function of academic capability, satisfaction, personal resourcing, personal health, and match of study programme to entry qualifications. From the point of view of funding against measures of retention or completion, it seems necessary to partial out those causes of non-completion which lie outside the institution's control - and it can be expected that the profile of causes will vary between institutions.

If the institution provides poor programmes or poor student support, then it is likely that students will seek to transfer out or leave higher education. An institution with a mission stressing commitment to the community might admit students who later turn out not to complete a programme of study. Students in the United Kingdom are typically encouraged to enrol for the highest award that they can (even if their intention is to leave with a lesser award) on the grounds that this is less expensive than to re-enrol for

progressively higher awards. Such a procedure may also be an inducement to a student to remain within a particular institution. When students do depart with a sub-degree award instead of a degree, this appears as 'non-completion' when set against the recorded initial intention.

Students may also leave their institution for positive reasons, such as obtaining a job, or to pursue their studies elsewhere.

Placement in employment

Whose interest?

Government has a natural interest in assessing the extent to which its investment on behalf of the public is paying off. Institutions, for a variety of reasons, want to be able to demonstrate that if students join them they have a good chance of success in the labour market. Intending students, for equally obvious reasons, are likely to favour an institution with a 'good' record in this respect, all other things being equal.

Validity

The Higher Education Statistics Agency [HESA] in the United Kingdom collects data on students' employment status some eight months after departure from higher education. The categories being used are as follows (HESA, 1995):

- full-time employment (paid or unpaid, including voluntary work);
- part-time employment (paid or unpaid, including voluntary work);
- self-employed;
- full-time further study or training;
- part-time further study or training;
- professional preparation time, e.g. portfolio preparation;
- looking for a job or course; and
- no other activity.

An indicator based on the position approximately eight months after leaving higher education is a dubious measure of success in obtaining employment. In today's constrained labour market it may take some time before a graduate or diplomate manages to obtain a post which is consistent with his or her qualifications. Better measures are employment data two or five years after leaving higher education, but these are subject to greater attenuation: the trade-off is between increased validity of the measure and the likelihood of reduced robustness of the data set. The data are also more expensive to collect. However, institutions in the United Kingdom are beginning to organise alumni associations along the lines of those that have in operation for a long time in the United States, and therefore the collection of such data can probably be assimilated into the alumni operation at a marginal cost.

The Linke Report showed how variable were employment data for fourteen Australian institutions, and commented on the relationship of this variability with both subject of study and regional economic circumstances. Its conclusions on the validity of employment data are worthy of quotation at some length.

> Whether at the aggregate level or by field of study there are serious problems in attempting to interpret institutional differences in graduate employment. There is clearly a need for a better understanding of the relative impact of regional economic, institutional, field of study and individual background factors on initial employment patterns before any meaningful interpretation could be made of institutional differences. And even then it would be necessary to monitor trends over time rather than rely on data from a single year. None of this, however, is to deny the potential value of employment data at the system level, where it has an obvious and essential role to play in both economic and educational planning . . . Rather it suggests that as yet little meaning can be attached to differential employment status of graduates in terms of specific institutional factors, and therefore that this data does not yet provide a useful performance indicator for institutional comparisons.
>
> (Linke, 1991 vol 1, p 89 )

Validity, for the Linke Committee, was distinctly problematic - a conclusion from which it is difficult to dissent.

Reliability

Higher education institutions in the UK are expected to reach a criterion return to HESA of 80% regarding the 'first destinations' of graduates and diplomates. Prior to the inception of the HESA system, a rather lower return was typically achieved. Some institutions are finding it difficult to reach the criterion rate, and are having to expend considerable effort in tracking down former students. Given that financial penalties are threatened for failure to reach the criterion, there is a risk that the reliability of the data may be compromised to a greater extent than is implied by the problematic validity of employment measures.

Side effects

Apart from the difficulty (noted in the preceding paragraph) of tracing former students, the side effects of using employment data are probably at present relatively low. However, if there is a move to fund institutions on the basis of outcomes of various kinds, then there is a risk that employment data will be given a 'gloss' which reflects well on institutions.

**Some other indicators**

Table I provides a brief commentary on a number of other indicators of institutional performance which have been selected (against the criterion of use or potential usability in higher education) from the plethora in the literature. No distinction has been made between management statistics and performance indicators on the grounds that what is a

management statistic at one level may be a performance indicator at another. An example here is that of the student/staff ratio, which is primarily a management statistic at the level of the institution, but which may be treated as a measure of institutional performance (efficiency) at a superordinate level.

| Indicator | Comment |
|---|---|
| Student/staff ratio | • Resourcing varies with subject area.<br>• Problem of counting research-oriented staff (especially if they are funded from external sources), and 'hybrid' staff, such as technician instructors: for some purposes, it is desirable to include them, for others it is not.<br>• Full-time equivalences, for both numerator and denominator, are not always easy to compute. |
| Cost per student | • Varies between subject areas.<br>• Depends on the way in which general institutional overheads are divided amongst institutional activities (e.g. teaching versus research). The balance of activities will vary between institutions (some are more heavily engaged in research than are others).<br>• Is affected by the conversion of student numbers into full-time equivalents. Flexibility in attendance is likely to introduce error.<br>• Is affected by whether the ratio refers to entering students or to those who succeed. |
| Research income per member of staff | • Varies substantially between subject areas, and is unstable year-on-year.<br>• As with student/staff ratio, which staff should be counted?<br>• How should a grant which is awarded for a number of years' work be counted - once, at the time of award, or divided between the years during which the work is undertaken?<br>• The actual spending against the grant may not match the stated level of the grant: which should be the indicator? |

| | |
|---|---|
| Research output per member of staff | • The first two points relating to research income also apply here<br>• The types of research output that are particularly valued vary with subject discipline*<br>• . . . and their value within disciplines is not necessarily consensual. |
| Library resources per student | • Statistics rarely cover the datedness and coverage of holdings.<br>• Nor do they generally deal with the tension between the provision of a shelf of standard texts and the need to provide a rich variety of sources.<br>• Other resources, particularly those based on developments in communications technology, are difficult to quantify. |
| Space allocation | • Normative figures may hide important detail, such as the appropriateness of the space for the intended activities, or the level of resourcing within the space (e.g. computer connections).<br>• Allowances may need to be made for the fact that buildings may not have been designed originally for educational use, and hence may not maximise the 'usable area' (e.g. if the corridor/room ratio is excessive). |

Table I Problems with some other indicators. The comments are illustrative and not exhaustive.

* See JPIWG (1994a, b) and Murphy et al (1994) for empirical evidence on this point.

**The problem of error**

It is apparent from the preceding sections that performance indicators are subject to error, some to a greater extent than others. At the level of the system, the amount of error may not matter very much, for errors in respect of individual institutions might roughly cancel out. The broad picture may simply not require fine brush strokes. Two examples are as follows.

1. The actual number of minority students may not matter very much at the level of a state system: all that the state may need to know is that participation from such students is significantly below the desired value, and that the evidence is sufficient for it to identify appropriate targets for the institutions under its aegis.

2. A system might be satisfied with the overall graduation rate (i.e. the return on its investment), and be little concerned by movement of students between institutions as they pursue their credits. On the other hand, if it were to transpire that there was a persistent flow of students away from a particular institution, then it might well wish to inquire as to the reasons why this has taken place.

As one moves from the system level 'downwards' to the level of the student experience, the indicators that are of importance change. They also tend to get 'softer'. Performance indicators relating to the student experience tend - inevitably - to be much more subjective (e.g. quality of teaching/learning, student satisfaction) than those used at the 'macro' level, such as funding per student, participation indexes and completion rates.

The preceding discussion of the selected indicators makes clear that each is surrounded by a penumbra of uncertainty whose magnitude is a function of a number of parameters. However, the size of the penumbra also depends on the use to which the indicator is being put. As noted earlier, errors of some magnitude may be unimportant in respect of general policy considerations, but highly important where inter-institutional comparisons are involved. If, say, there is a ten per cent error in a participation statistic at the level of the system, this may be insignificant in respect of system needs but, if such an error were replicated at the level of institutions, some might be disadvantaged unfairly in respect of a remedial funding initiative. Dropping down a level, an institution might be reasonably satisfied with the overall satisfaction rating given by its students to their learning experiences, but might unfairly castigate a department for appearing to have a below-norm level of satisfaction when the actual figure was influenced by external factors (e.g. the requirement of a professional body that the curriculum be knowledge-intensive rather than interactive).

Scientists are well-used to estimating the level of error associated with their findings. Although the errors associated with performance measures are probably more difficult to estimate, it should be possible to give these an order of magnitude and therefore to produce tolerances (akin to confidence limits, but necessarily less sharply defined) within which performances are adjudged acceptable. Fuzzy logicians would have no difficulty with fuzzy measures, and with fuzzy process control. Fuzzy process control is probably the best that can be achieved with such a complex matter as education, whatever the level of engagement with it.

**Some policy considerations**

Can performance indicators be trusted? The preceding discussion suggests that some are more trustworthy than others, and that trustworthiness is a function of the level at which they are being used. None of the selected indicators is free from uncertainty, and the same is likely to be true of others. Some indicators will be more meaningful than others to a lay audience. It seems likely that most indicators are not susceptible of refinement to the level of accuracy that a researcher would desire - at least not without producing a plethora of qualifications which might obscure the essential 'message'. Extra expenditure on refining indicators and on collecting extra detail may not be cost-beneficial, because policy-makers (at any level) typically have to make do with rough and ready information: for their purposes, 80 per cent accuracy today is often preferable to 95 per cent tomorrow. As Ewell and Jones (1994, p16) put it:

> Many promising indicator systems fail simply because they are too expensive, too complex, too time-consuming, or too politically costly to implement. Often the simplest is the best, even if it initially seems less technically attractive.

The corollary of Ewell and Jones's point is the obvious one - that indicators are likely to indicate performances in a relatively rough and ready way, rather than to be precise measures. The danger, for users and interpreters, lies in focusing on the mean whilst overlooking the standard deviation, to strike an analogy with statistics.

Given the likely imprecision in performance indicators, they are particularly vulnerable to partiality in their use: this reiterates a point made in the Introduction. Vested interest may give an indicator a colouring not warranted by the data, and each

use of an indicator has to be understood with reference to the user's likely purposes. Just as one has to be cautious about trusting indicators, so one has to be cautious about trusting the users of indicators.

Probably the majority of uses of indicators reflect both accountability and enhancement, and it has been argued that is preferable to separate the two 'agendas'. Ewell (1994, p162) quotes an agency staff member in Kentucky as saying "You can miss the point and get real fuzzy if you try to do everything together". The primary raison d'être of policy-makers is to maximise the effectiveness of the resources at their disposal, which implies an enhancement orientation without denying the need for procedures designed to fulfil the expectations of accountability. It seems likely that the Kentuckian's 'real fuzziness' reflects a more general failure fully to appreciate the inherent duality of indicators (in an earlier article [Yorke, 1995] I described them for this reason as 'siamese twins') and to be aware of when the indicators were being used primarily for enhancement and when primarily for accountability. The Kentuckian's fuzziness might also reflect a failure to appreciate the political arena in which indicators were being used.

If indicators - as has been argued above - are likely to provide information of varying degrees of fuzziness, then a corollary is that their interpretation is likely to be less reliable the further the interpreter is from the source of the indicator. For many - but not all - situations, a construal of higher education in terms of a 'nested set' of levels may be helpful (see Yorke, 1996a). Using this perspective, indicator data are evaluated and acted upon at the 'lowest' level possible, and 'higher' levels are

expected to audit whether the data have been obtained and acted upon in an appropriate manner. Some matters are capable of being dealt with by individual members of staff (e.g. feedback on a set of teaching sessions), whereas others have to be referred upwards (such as the provision of extra computer workstations).

Inspectorial activities, such as teaching quality assessment currently run by the funding councils in the United Kingdom, ought on this approach to be used as cross-checks that institutions are doing what they should be doing rather than be used for saturation coverage of teaching quality. This kind of approach has a lot of features in common with the notion of total quality management [TQM] - but a variant of TQM that is developed in order to take account of the lack of precision (relative to the production of industrial artefacts) with which educational activities typically are conducted (see Yorke, 1997).

**Coda**

The belief that performance indicators can provide accurate information on institutional functioning appears to have weakened in recent years as early optimism has become tempered by debate and experience: a case in point is that of the Higher Education Funding Council for England which set out on its programme of teaching quality assessment by claiming that six indicators would help it do determine whether quality was excellent, satisfactory or unsatisfactory, but subsequently downgraded these indicators in both number and application to mere 'statistical indicators' (compare HEFCE, 1992, with HEFCE, 1994).

That the initial expectations regarding performance indicators have not yet been fulfilled is revealed in the work of Cave et al (1991; 1997). In the second edition of their book on performance indicators Cave et al (1991, p24) defined an indicator (rather blandly, it has to be said) as

> 'an authoritative measure - usually in quantitative form - of an attribute of the activity of a higher education institution.'

The definition in the third edition is similar, but has one notable omission - the word 'authoritative'.

**References**

Astin AW, Tsui L and Avalos J (1996) **Degree attainment rates at American colleges and universities: effects of race, gender and institutional type**. Los Angeles: Higher Education Research Institute, University of California.

CAPIOA (1993) **Report of the Committee on Accountability, Performance Indicators and Outcomes Assessment to the Minister's Task Force on university accountability**. Toronto, Ontario: CAPIOA.

Cave M, Hanney S and Kogan M (1991) **The use of performance indicators in higher education: a critical analysis of developing practice** (2nd edition). London: Jessica Kingsley.

Cave M, Hanney S, Henkel M and Kogan M (1997) **The use of performance indicators in higher education: a critical analysis of developing practice** (3rd edition). London: Jessica Kingsley.

Chapman K (1994) Variability of degree results in geography in United Kingdom Universities, 1973-1990: preliminary results and policy implications. **Studies in Higher Education 19,** pp 89-102.

CNAA/PCFC (1990) **The measurement of value added in higher education**. London: Council for National Academic Awards and the Polytechnics and Colleges Funding Council.

Dobson I, Sharma R and Hayden A (1996) **Evaluation of the relative performance of commencing undergraduate students in Australian universities**. Adelaide: Australian Credit Transfer Agency.

Elton L (1998) Are UK degree standards going up, down or sideways? **Studies in Higher Education 23** (forthcoming).

Ewell PT (1994) Developing statewide performance indicators for higher education: policy themes and variations. In Ruppert SS (ed) **Charting higher education accountability: a sourcebook on state-level performance indicators**. Denver, CO: Education Commission of the States, pp 147-166.

Ewell PT and Jones DP (1994) Pointing the way: indiocators as policy tools in higher education. In Ruppert SS (ed) **Charting higher education accountability: a sourcebook on state-level performance indicators**. Denver, CO: Education Commission of the States, pp 6-16.

Folger J and Jones DP (1993) **Using fiscal policy to achieve state education goals**. Denver, CO: Education Commission of the States.

Gaither G, Nedwek B and Neal JE (1994) **Measuring up: the promises and pitfalls of performance indicators in higher education** (ASHE-ERIC Report #5). Washington, DC: George Washington University.

Green D, with Brannigan C, Mazelan P and Giles L (1994) Measuring student satisfaction: a method for improving the quality of the student experience? In Haselgrove S (ed) **The student experience**. Buckingham: SRHE and Open University Press, pp 100-107.

HEFCE (1992) **Quality assessment** (Consultation document). Bristol: Higher Education Funding Council for England.

HEFCE (1994) **The quality assessment method from April 1995**. Bristol: Higher Education Funding Council for England.

HEQC (1996) **Inter-institutional variability of degree results: an analysis in selected subjects**. London: HEQC.

HEQC (1997) **Graduate Standards Programme: Final Report**. London: HEQC.

HESA (1995) **Student first destination supplement coding manual, 1994-1996**. Cheltenham: Higher Education Statistics Agency.

Johnes J and Taylor J (1990) **Performance indicators in higher education**. Buckingham: SRHE and Open University Press.

JPG (1996) **Joint Planning Group for Quality Assurance in Higher Education: Final Report** (17 December). London: CVCP.

JPIWG (1994a) **Management statistics and performance indicators in higher education** (Consultative Report of the Joint Performance Indicators Working Group). Bristol: Higher Education Funding Council for England.

JPIWG (1994b) **Explanatory and statistical material to accompany Consultative Report**. Bristol: Higher Education Funding Council for England.

Kells HR, ed (1993) **The development of performance indicators in higher education: a compendium for twelve countries**. Paris: OECD.

Linke R, Chair (1991) **Performance indicators in higher education** (Report of a trial evaluation study commissioned by the Commonwealth Department of Employment and Training. 2 vols). Canberra: Australian Government Publishing Service.

NCIHE (1997) **Higher education in the learning society** [Report of the National Committee of Inquiry into Higher Education]. London: Her Majesty's Stationery Office.

O'Leary J, ed (1996) **The Times good university guide 1996**. London: Times Books

Ramsden P (1991a) A performance indicator of teaching quality in higher education: the Course Experience Questionnaire. **Studies in Higher Education 16**, pp 129-150.

Ramsden P (1991b) Report on the Course Experience Questionnaire trial. In Linke R, Chair, *op cit*, Volume 2, pp1-85.

Ruppert SS, ed (1994) **Charting higher education accountability: a sourcebook on state-level performance indicators**. Denver, CO: Education Commission of the States.

SCFC (1992) **Performance indicators** [Final Report of the Working Party of the SCFC]. Edinburgh: Scottish Centrally Funded Colleges.

Stecklow S (1995) Cheat sheets: colleges inflate SATs and graduate rates in popular guidebooks. **The Wall Street Journal** (Western edition), 27 April, p A16.

Turner D and Woolf H, with Bridges P, Bourdillon B, Cooper A, Fox W, Haines C and Yorke M (1998) Honours classifications: the need for transparency. **The New Academic** (forthcoming).

Williams G (1994) Achieving quality in higher education: a student-centred approach to self-evaluation. Newport: Gwent College of higher Education (mimeo).

Yorke M (1995) Siamese twins? Performance indicators in the service of accountability and enhancement. **Quality in Higher Education 1** (1), pp 13-30.

Yorke M (1996a) **Indicators of programme quality**. London: HEQC.

Yorke M (1996b) Shouldn't quality be enhanced, rather than assessed? **Tertiary Education and Management 2** (1), pp 86-94.

Yorke M, Cooper A, Fox W, Haines C, McHugh P,Turner D and Woolf H (1996) Module mark distributions in eight subject areas and some issues they raise. In Jackson N (ed) **Modular higher education in the UK in focus**. London: HEQC, pp105-107.

Yorke M (1997) The elusive quarry: total quality in higher education. **Tertiary Education and Management 3** (2), pp145-156.

Yorke M, with Bell R, Dove A, Haslam L, Hughes Jones H, Longden B, O'Connell C, Typuszak R and Ward J (1997) **Undergraduate non-completion in England**. Final report of a research project commissioned by the Higher Education Funding Council for England.